

Study of Item Nonresponse for a Sample of Questionnaires from the 2007 Commodity Flow Survey (CFS)

Hossain Eftekhari-Sanjani

Bureau of Transportation Statistics, Research and Innovative Technology Administration, U.S. Department of Transportation, 1200 New Jersey Avenue S.E.
Washington, D.C. 20590

Abstract

The CFS is a quinquennial, nationwide survey that collects data on the shipment of goods within the United States. It is a component of the Economic Census, and is conducted jointly by the U.S. Department of Transportation's Research and Innovative Technology Administration, Bureau of Transportation Statistics, and the U.S. Department of Commerce's Census Bureau. A sample of the 2007 CFS completed questionnaires was selected for in-depth investigation of missing or obviously bad responses to items. The CFS questionnaire is complex, and prior to implementation of the 2007 survey, extensive efforts were undertaken to improve the flow and comprehension of the questionnaire. The research presented in this paper was aimed at assessing existing item nonresponse, the type of items most affected, and the relationship of item nonresponse to the job title of the person completing the form. Results indicated that about 70% of respondents are in management occupations, those in transportation-related occupations had the lowest levels of item nonresponse, and items which require consulting establishment's records had higher rates of item nonresponse, and item nonresponse varied based on industry classification of establishments.

Key Words: Item Nonresponse, Commodity Flow Survey, Transportation, Establishment Survey, Data Quality

1. Introduction

The importance of unbiased data in survey research becomes apparent when one considers that, nationally, policy decisions in health, transportation, education, and other elements of the nation's economy are often based on survey data. Survey nonresponse is one of the most important issues affecting the quality of survey data and can lead to serious bias in the survey results (Dixon 2005; Rassler and Riphahn 2006). Survey nonresponse is of two types, unit and item nonresponse (Elliot et. al. 2005). Unit nonresponse happens when a respondent fails to respond to the entire questionnaire as a whole. In contrast, item nonresponse occurs when a respondent provides some, but not all the information requested on the questionnaire, or when the responses to some of the questionnaire items are not usable.

Studies in survey item nonresponse yield several perspectives on reasons for, and factors affecting, item nonresponse and offer numerous alternative approaches for dealing with missing data values. The literature suggests that item nonresponse can be attributed to both item and respondent characteristics (Rassler and Riphahn 2006; Elliot et. al. 2005; Dixon 2005). Among the reasons mentioned for not providing the data are: difficulty in recalling information for the response, sensitivity of information requested, difficulty in accessing the information needed to respond, the complexity of providing a response, and awkward question format. In one study of the interaction of item and respondent characteristics, questions perceived as more important by respondents had less item nonresponse, and higher-educated respondents had more item nonresponse, particularly with regard to women and Hispanics (Dixon 2005). On the other hand, Dixon 2005, found that education lowered the odds of nonresponse, and age slightly increased them. Common approaches to dealing with survey item nonresponse include, excluding the observations with missing data items, imputing missing values to generate a complete set of data, weighting the data from some respondents in subsamples to compensate for missing values in other subsamples, and model-based procedures in which predictions of the missing values are calculated. The most commonly used approach, excluding incomplete observations, can lead to substantially biased survey results (Rassler and Riphahn 2006).

1.1 Establishment Survey Item Nonresponse

Establishment surveys, such as the 2007 Commodity Flow Survey, are often large-scale studies that collect data from organizations, businesses and institutions, and provide key data for government statistics and policy makers. Several reasons for item nonresponse in establishment surveys include: the complexity of the questionnaires, unclear instructions, amount and type of the data sought, excessive cost of retrieving the required records, lack of knowledge of the responding person of information requested and his/her authority in obtaining required records, availability of the required records at the contacted establishment, and the sensitivity of confidential data (Fisher, et.al. 2003; Slowinski 1988,). Sometimes the information sought by an establishment survey is maintained in different locations of an organization, and thus, input from various individuals in the establishment may be required to respond to various items of the questionnaire. Additionally, large establishments that are included in most of the establishment surveys are overwhelmed by surveys, and smaller establishments with limited resources may tend to answer the items for which the answers are readily available and skip the other items. Edwards (1994) studied the relationship between the respondents' level and functional role in the organization and item nonresponse, and suggested a relationship between respondents' functional role and item nonresponse in some cases. He further suggests that "the best survey respondent" in an establishment survey is someone who is familiar with the system, knows how to locate the requested information, and has the authority to access the information and complete the survey.

2. Background of the CFS

The Commodity Flow Survey (CFS) is a joint effort between the U.S. Census Bureau and the Research and Innovative Technology Administration's Bureau of Transportation Statistics, U.S. Department of Transportation, and is a component of the quinquennial economic census. The survey is a mail-out/mail-back sample survey of business establishments in mining, manufacturing, wholesale, and other selected industries, and provides key statistics about the transportation of freight in the United States. It is the only source of nationwide data on the movement of goods from origin to destination by all modes of transportation and for multi-modal combinations.

The CFS uses a complex, three-stage sample design (i.e., establishments, reporting weeks, shipments) that is stratified by geography, industry, and the establishment size. The sample size for the 2007 CFS was over 100,000 establishments. A single questionnaire was used for all establishments, and it was designed to collect information that can be very heterogeneous given that the industries vary in type and size of enterprise, volume of outbound shipments, and types of commodities shipped. Additionally, respondents are asked to select a systematic random sample of their establishment's total number of outbound shipments during a specified week using the instructions provided on the questionnaire and in the accompanying instruction material. Each establishment selected into the 2007 CFS sample was mailed a questionnaire four times during the calendar year and asked to report on a sample of individual shipments during a one-week period in each calendar quarter.

Completion of the questionnaire primarily consisted of three activities whereby respondents were asked to: (1) verify the name and address of the establishment printed on the questionnaire; (2) provide the total number of outbound shipments' for a given week and draw a systematic random sample from the total shipments for that week; and (3) provide several items of information about each of the sampled shipments they selected. This information included:

- date on which the shipment was made
- the type of commodity shipped (along with corresponding classification code);
- value and weight of the shipment;
- mode(s) of transportation used
- domestic destination
- indication of whether the shipment was an export
- *if an export*, the foreign destination and U.S. (domestic) port of exit
- indication of whether the shipment was a hazardous material
- *if a hazardous material*, corresponding classification code.

In planning for the 2007 CFS, extensive efforts were undertaken to improve the quality of the information requested and ease the burden of the respondent in providing these data. One specific activity that was implemented to assist with improving item response was a multi-stage cognitive interviewing effort. A critical part of the testing involved face-to-face interviews with company representatives to better understand the steps they take to complete the questionnaire, problems they encounter in completing the questionnaire, and their interpretation of instructions on the questionnaire and the accompanying materials. In total, the questionnaire underwent three rounds of cognitive testing with a total of 75 participating establishments. After each of the first two rounds of testing, the questionnaire was modified based on the findings of the previous round and retested to check for improvement.

In spite of these and other efforts, a significant amount of item nonresponse still occurred in the conduct of the 2007 CFS. To better understand the types and causes of CFS item nonresponse, a special effort was initiated after the data collection phase of the survey. Prior to this study, no attempts had been made to select a systematic sample of the returned questionnaires from any of the prior Commodity Flow Surveys and literally examine each of them (prior to any callbacks, editing, or imputation), for the purpose of identifying any possible patterns of item nonresponse, and to check for any written comments that might have been provided on the questionnaire. Moreover, despite the fact that some information requested in the CFS survey can be sensitive for some establishments, and access to, and knowledge of the information may be restricted to certain positions in the establishment, the relationship between nonresponse and position title of the responder had not yet been studied in the CFS

3. Objectives of the Study

The objectives of this study were to obtain information from the questionnaires beyond the data entered for the production of the CFS data tables or summarized in edits, and to augment our prior findings from the CFS questionnaire improvement processes. Specifically, the objectives of this study were to:

- investigate possible patterns of item nonresponse.
- compare item nonresponse across various parts of the questionnaire.
- identify the job title of the person who completed the questionnaire, and examine relationship between job titles and item nonresponse.
- examine possible differences in item nonresponse based on the industry classification of establishments.

4. Methods

Each establishment in the 2007 CFS sample was mailed four questionnaires, one questionnaire during each calendar quarter. The frame for the sample of this study included about 69,000 establishments in the 2007 CFS sample that had returned at least one of the four questionnaires they had received by December 2007. A random sample of 870 establishments was selected for this study. The selected establishments, in total had returned 2,875 completed questionnaires in total.

An electronic image of each 2007 CFS returned questionnaire was taken and saved in a database maintained by the U.S. Census Bureau. Authorized BTS analysts were able to access each image, review its contents, and enter the desired data on a specially developed spreadsheet interface. The spreadsheet included a list of the issues that each analyst needed to check for, and clearly identified cells for entering requested data. To facilitate the data entry process and reduce chances of data entry errors, each cell of the spreadsheet contained pull down instructions and definitions for codes to be entered in the cell, and accepted only one of the valid codes for that cell.

5. Results

A total of 2,875 questionnaires from 870 establishments were examined for this study. Table 1 includes the distribution of the returned questionnaires with regard to the calendar quarter. Since the 2007 CFS fourth quarter questionnaires were mailed shortly before or during the fourth calendar quarter, not all respondents had returned their questionnaires by mid- December. As a result, the sample did not include the very late responders in Quarter 4.

Table 1. Distribution of returned questionnaires by each quarter

<i>Returned Questionnaires from the Establishments by Quarter (Q1-Q4)</i>	<i>Sample of this study</i>		<i>Population for this Study</i>	
	<i>Count</i>	<i>Percent</i>	<i>Count</i>	<i>Percent</i>
Q1 only	53	6.1	5,433	7.9
Q2 only	26	3.0	2,697	3.9
Q3 only	12	1.4	2,171	3.2
<i>Completed Q4 only</i>	<i>0</i>	<i>0.0</i>	<i>1,583</i>	<i>2.3</i>
Q 1 and Q2	39	4.5	2,685	3.9
Q1 and Q3	21	2.4	1,422	2.1
Q1 and Q4	10	1.1	1,294	1.9
Q2 and Q3	11	1.3	1,320	1.9
Quarters 2 and 4	6	0.7	942	1.4
Q 3 and Q4	11	1.3	2,108	3.1
Quarters 1 & 2 & 3	47	5.4	3,850	5.6
Q1 & Q2 & Q4	24	2.8	1,825	2.7
Q1 & Q3 & Q4	33	3.8	2,965	4.3
Q2 & Q3 & Q4	32	3.7	3,924	5.7
All Quarters	545	62.6	34,628	50.3
Total establishments	870	100.0	68,847	100.0

5.1 Group I Items: General Information about Establishment and Respondent

Questionnaire items in the 2007 CFS can be classified into two broad groups of questions. Group I includes eleven items of general information about the establishment and the person responding to the survey. These items request verification of the name, address, and the operational status of establishment, name telephone number and the job title of the person completing the form, total number of establishment's outbound shipments, and total value of shipments made during a one-month period. Generally, except for the latter two technical items, responding to the rest of the items in this group should not require reference to documentation.

The overall rate of nonresponse for the eleven items in Group I was 2.5%. That is, out of all the expected responses for these eleven items (11 x 2,875), only 792 (2.5%) were missing. The nonresponse rate calculated the same way for nine of these eleven items that collected general address and contact information was 2.1%, while the nonresponse for the two technical items inquiring the total number of outbound shipments and total value of a recent month's shipments was 4.3%.

The rate of item nonresponse per questionnaire for all Group I items is presented in Table 2. Over eighty two percent (82.5%) of respondents had answered all eleven items in Group I. Almost twelve percent (11.9%) had not responded to only one of the eleven items. Out of the eleven items, 2.8%, and 1.7% left two and three items unanswered, respectively. Only about one percent had left between four to seven items unanswered. In Group I items, 86.5% had answered all of the nine nontechnical items. Almost ten percent (9.8%) left only one, and 2.5% left two of the nontechnical items unanswered. Less than one percent (0.9%) left three, and 0.3% left four to five items unanswered. Regarding the two technical items in Group I, 92.3% had responded to both, and 6.7% had responded to only one of them. One percent of respondents left both of the technical items unanswered.

Table 2. Distribution of the cumulative count of item nonresponse for the general items in Group I.

Cumulative count of item nonresponse	All Group I Items (Total of 11 items)		Non-technical Items (Total of 9 items)		Technical Items (Total of 2 items)	
	Count	Percent	Count	Percent	Count	Percent
Responded to all items	2,373	82.5	2,486	86.5	2,654	92.3
One item unanswered	343	11.9	282	9.8	193	6.7
Two items unanswered	81	2.8	72	2.5	28	1.0
Three items unanswered	49	1.7	27	0.9	N/A	N/A
Four or five items unanswered	22	0.8	8	0.3	N/A	N/A
Six or seven items unanswered	7	0.2	N/A	N/A	N/A	N/A
Total	2,875	100	2,875	100	2,875	100

5.2 Group II Items: Detailed Information about Characteristics of Shipment

The second group of questions (Group II) in the 2007 CFS questionnaire sought detailed information about the characteristics of a sample of the establishment's outbound shipments, made during a specified week. While all respondents were required to respond to all eleven items in Group I, the number of items in Group II depended on the number and the type of shipments reported by the respondents. The number of reported shipments varied between zero and forty and averaged around 25. For each reported shipment, respondents were asked to provide between thirteen to seventeen data items that require access to operational and financial records of the establishments. The number of required items for each shipment is based on the type of the commodity shipped, the mode(s) of transportation used, and the shipment's foreign or domestic destination.

There are 13 items for which responses are required for all of the reported shipments. To compute an overall rate of nonresponse for these items, the sum of the nonresponse to these items for all the reported shipments on each questionnaire, was divided by the product of 13 and the number of shipments reported on the questionnaire. Nonresponse analysis of the data for these items indicated that almost 67 percent of the respondents had no item nonresponse. For 10% of the respondents the rate of item nonresponse for all reported shipments combined was more than zero, but less than one percent. This cumulative nonresponse rate was between one to less than five percent for 8% of respondents. Ten percent of respondents had a nonresponse rate of between five to less than ten percent for all the shipments they reported, combined. The remainder four percent had a nonresponse rate of between 10 to 33 percent for all shipment characteristic items combined. The rate of cumulative item nonresponse for all the thirteen items required for all reported shipment is summarized in Figure 1.

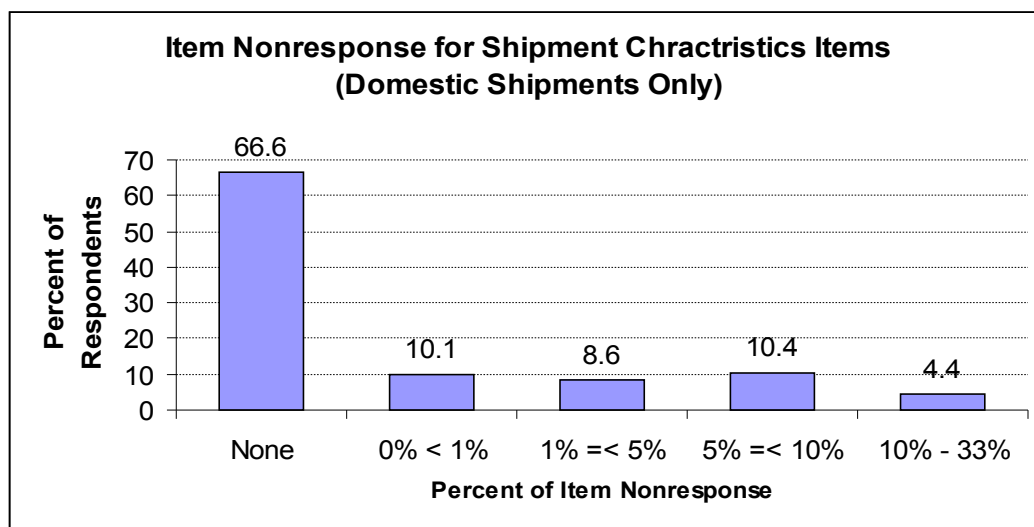


Figure 1: Percentage of Item Nonresponse of Shipment Characteristics Items.

The rate of item nonresponse for eight of the most important items required for each reported shipment is presented in Figure 2.



Figure 2: Item Nonresponse for Eight of the More Important Shipment Related Items.

5.3 Occupational Classification of the Respondents

Every returned questionnaire included the name and the job title of the individual who had responded to the survey. There were a total of 450 unique job titles reported on the 2,875 returned questionnaires. The Bureau of Labor Statistics’ Standard Occupational Classifications (SOC) major and minor groups were used to classify these 450 self-reported job titles. Figure 2 depicts the percentages of the respondent within the major and minor groups of SOC. The majority of respondents (69%) were in the management occupations. This group included General and Operations Managers (27%); Financial Managers (26%); Transportation, Storage, and Distribution Managers (8%); and other management occupations (8%). Seven percent (7%) were accountants and auditors, and three percent (3%) were in other business and financial operations occupations. First-line supervisors or managers of office and administrative support workers comprised seven percent (7%) of the respondents, and an equal percentage of the respondents were in other administrative and support occupations. Only four percent (4%) of the questionnaires were completed by non-managers in transportation and material moving occupations. Three percent (3%) of the respondents had other occupations.

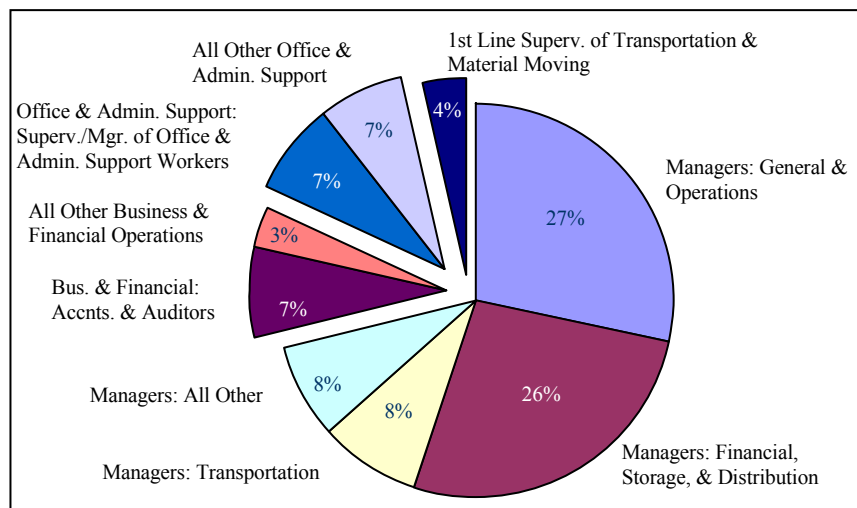


Figure 3: Distribution of Job Titles of Respondents within SOC Groups

5.4 Item Nonresponse by Industry Classification of the Establishments

The establishments in the sample were classified by their North American Industrial Classification System (NAICS). The means and standard errors of the percent nonresponse for the sum of items in Group II (shipment related items that are required from all respondents) are presented in Table 3. One-way analysis of variance indicated significant differences ($p=0.000$) in the percent of item nonresponse among various NACS industries. Using the Dunnett’C post Hoc test, percent of item nonresponse significantly differed among manufacturing (NAICS 31-33), wholesale and retail (NAICS 42 & 45), and information (NAICS 51) groups.

TABLE 3. PERCENT NONRESPONSE FOR SHIPMENT CHARACTERISTICS ITEMS WHICH ARE REQUIRED FOR ALL REPORTED SHIPMENTS.

NAICS	Industry	N	Mean	S. E.	Item Nonresponse	
					Min.	Max.
21	Mining	95	2.82	0.53	0.0	23.6
31 - 33	Manufacturing	1,352	2.87 *	0.16	0.0	40.0
42 & 45	Wholesale and Retail	1,159	3.93 *	0.20	0.0	45.7
49	Transportation & warehousing	57	2.71 *	0.69	0.0	26.9
51	Information	24	0.76	0.58	0.0	13.3
55	Mgmt of companies & businesses	53	4.09	0.96	0.0	26.7
Total	Total	2,740	3.32	0.12	0.0	45.7

*Significantly different ($p<0.001$) from each other.

5.5 Discussion

Results of the study confirm lower nonresponse rates for general items, than for technical items where answering the item requires consulting the establishment’s operational and financial documents. Preliminary analysis of data reveals some possible differences in the percentage of item nonresponse based on the occupational title of the respondent. Analysis, not reported in this paper due to space limitations, indicated that the percentage of the respondents in transportation occupations is positively correlated with an estimate of the size of establishment (total value of shipments for the most recently completed month). As the resources would permit, the results of this study will be augmented by adding the data from some additional questionnaires that have been selected from universe of the very late responders to 2007 CFS, and by the further analysis of the final data.

Acknowledgements

The author would like to thank Mohamad Adhi Dipo, Chester Ford, and Getachew Mekonnen for their diligent efforts in reviewing the questionnaires, and recording the data. Thanks are also due to Joy Sharp, Ronald Duych, Michael Margreta, and Alan Jeeves for their reviews and invaluable comments and suggestions.

References

- Dixon, J. (2005). *Comparison of item and unit nonresponse in household surveys*. Washington DC. Bureau of Labor Statistics. Retrieved July 2, 2008 from <http://www.bls.gov/osmr/abstract/st/st050080.htm>
- Edwards, W.S. (1995). *The effect of respondent level and function on item nonresponse in the 1994 NEHIS*. Paper presented at the 1995 Annual Meeting of the American Statistical Association. Retrieved from http://www.amstat.org/sections/SRMS/proceedings/papers/1995_030.pdf

- Elliott, M., Edwards, C. Angeles, J., Hambarsoomians, K., and Hays, R. (2005). Patterns of unit and item nonresponse in the CAHPS® hospital survey. [Electronic version] *Health Services Research* 40 (6) December 2096-2119. Health Research and Educational Trust doi: 10.1111/j.1475-6773.2005.00476.x
- Fisher, S., Bosley, J., Goldenberg, K., Mockovak, W., Tucker, C., (2003). *A qualitative study of nonresponse factors affecting BLS establishment surveys*. Washington DC. Bureau of Labor Statistics. Retrieved July 2, 2008 from <http://www.bls.gov/osmr/abstract/st/st030230.htm>
- Rassler, S. and Riphahn, R. (2006). Survey item nonresponse and its treatment. [Electronic version] *Allgemeines Statistisches Archiv*, 90 (1) March, 217-232. Retrieved from Ingenta Connect <http://www.ingentaconnect.com/>
- Rosen, R., Clayton, R., Rubino, T., Jr., (1991). *Controlling nonresponse in an establishment survey*. Bureau of Labor Statistics. Paper presented at the 1991 Annual Meeting of the American Statistical Association. Retrieved from http://www.amstat.org/sections/SRMS/proceedings/papers/1991_100.pdf
- Slowinski, S.M. (1988). *Control and measurement of nonresponse error in establishment surveys*. Federal Reserve System. Paper presented at the 1988 Annual Meeting of the American Statistical Association. Retrieved from http://www.amstat.org/Sections/Srms/Proceedings/papers/1988_058.pdf
- Smith, P. (2004). *Perspective on response rates and nonresponse in establishment surveys*. Paper presented at December 14, 2004 meeting of the United States Federal Economic Statistics Advisory Committee, Washington, DC. Retrieved on August 21, 2008 from <http://www.bea.gov/about/pdf/PerspectiveonResponseRatesandNonresponseinEstablishmentSurveys.pdf>
- Tersine, A. (1998). *Item nonresponse: 1996 American Community Survey*. Washington, DC: U.S. Census Bureau. Retrieved June 28, 2008, from <http://www.census.gov/acs/www/AdvMeth/Papers/ACS/Paper13.htm>
- Wood, M. (2005) *A cumulative count, general approach to item-nonresponse* Paper presented at the annual meeting of the American Association For Public Opinion Research. Retrieved from <http://www.amstat.org/Sections/Srms/Proceedings/y2005/Files/JSM2005-000857.pdf>

¹ The views expressed in this paper are attributable only to the authors and are not necessarily those of BTS or of the Bureau of the Census.